

Практикум 12

Задание 1: частоты старт кодонов

Таблица 1. Частоты встреченных старт-кодонов

Escherichia coli	Candidatus Gracilibacteria bacterium	Mycoplasma pneumoniae
ATG 3890 GTG 338 TTG 80 ATT 4 CTG 2 TTC 1	ATG 1129 GTG 41 TTG 23 ACA 1 TCA 1 TCT 1	ATG 629 GTG 60 TTG 53 ATT 8 ATA 4 TTA 3 CAA 2 CTC 2 AAA 1 ACT 1 ATC 1 CTG 1 GAA 1 GGA 1 GTT 1 TCT 1

Исходя из результатов, можно заметить наличие альтернативных старт-кодонов у исследуемых объектов. Однако, не смотря на разнообразие старт-кодонов, наиболее часто встречаемым у 3х бактерий является канонический кодон ATG.

Замена канонического кодона на альтернативные распространено среди прокариотических организмов. Однако альтернативные кодоны в начале открытой рамки считывания в некоторых случаях заметно влияют на эффективность процессов реализации генетического материала. Однако были описаны и случаи сопоставимой эффективности альтернативных кодонов [1].

Кроме того можно отметить довольно большое разнообразие старт кодонов у организма *Mycoplasma pneumoniae*. Такое явление может быть связано с образом жизни данного организма. *Mycoplasma pneumoniae* ведет паразитический образ жизни, является возбудителем бактериальной пневмонии. В связи с этим *Mycoplasma pneumoniae* обладает рядом особенностей связанных с паразитическим образом жизни [2]. Можно предположить, что некоторые особенности присутствуют и на уровне генома. Так в процессе приспособления к паразитическому образу жизни у организма могут обнаруживаться последовательности, на которые не действует естественный отбор. Крайней степенью таких преобразований в геноме могут являются псевдогены.

Перечисленные выше последовательности в геноме и могут иметь неклассические старт-кодоны.

Механизмов возникновения таких старт кодонов несколько. Можно заметить, что часто встречаются альтернативные старт кодоны, отличающиеся от каноничного всего на 1 нуклеотид, что говорит о высокой вероятности возникновения единичной мутации. В таких случаях имеет место реализации воббл-механизма.

Другим возможным путем формирования альтернативных старт кодонов может быть сдвиг рамки считывания.

Задание 2: необычные стоп-кодоны E. Coli

Таблица 2. Название последовательностей, в которых стоп кодоны встречаются не только в конце последовательности

Название последовательности с описанием	Ссылка на описание региона на сайте NCBI
lcl U00096.3_cds_b4587_250 [gene=insN] [locus_tag=b4587] [db_xref=ASAP:ABE-0285253,ECOCYC:G6130] [protein=IS911A regulator fragment] [pseudo=true] [location=join(270278..270540,271764..272190)] [gbkey=CDS]	https://www.ncbi.nlm.nih.gov/nuccore/U00096.3?location=270278:270540,271764:272190
lcl U00096.3_cds_AAD13438.1_1459 [gene=fdnG] [locus_tag=b1474] [db_xref=UniProtKB/Swiss-Prot:P24183] [protein=formate dehydrogenase N subunit alpha] [transl_except=(pos:586..588,aa:Sec)] [protein_id=AAD13438.1] [location=1547401..1550448] [gbkey=CDS]	https://www.ncbi.nlm.nih.gov/nuccore/U00096.3?location=270278:270540,271764:272190#comment_U00096.3
lcl U00096.3_cds_AAD13456.1_3824 [gene=fdoG] [locus_tag=b3894] [db_xref=UniProtKB/Swiss-Prot:P32176] [protein=formate dehydrogenase O subunit alpha] [transl_except=(pos:586..588,aa:Sec)] [protein_id=AAD13456.1] [location=complement(4082772..4085822)] [gbkey=CDS]	https://www.ncbi.nlm.nih.gov/nuccore/U00096.3?location=270278:270540,271764:272190#comment_U00096.3

lcl U00096.3_cds_AAD13462.1_3997 [gene=fdhF] [locus_tag=b4079] [db_xref=UniProtKB/Swiss-Prot:P07658] [protein=formate dehydrogenase H] [transl_except=(pos:418..420,aa:Sec)] [protein_id=AAD13462.1] [location=complement(4297219..4299366)] [gbkey=CDS]	https://www.ncbi.nlm.nih.gov/nuccore/U00096.3?location=270278:270540,271764:272190#comment_U00096.3
---	---

Среди кодирующих последовательностей организма *Escherichia coli* str. K-12 substr. MG1655 были обнаружены 4 последовательности с нетерминирующими стоп-кодонами.

1. Согласно описанию последовательность является псевдогеном. Кроме того в таблице особенностей последовательности можно обнаружить информацию о наличии мобильного генетического элемента (<264..690 /mobile_element_type="insertion sequence:IS911A-1")
Поэтому на эту последовательность возможно не действует естественный отбор и стоп-кодон сохраняется, не уменьшая эффективность реализации генетической информации.
2. Последовательности 2-4 согласно описанию содержат кодоны, кодирующие аминокислоту селеноцистеин: [transl_except=(pos:586..588,aa:Sec)], [transl_except=(pos:586..588,aa:Sec)], [transl_except=(pos:586..588,aa:Sec)], [transl_except=(pos:418..420,aa:Sec)]
Селеноцистеин кодируется кодоном соответствующим стоп кодону [3].

Задание 3: Стоп кодоны

Таблица 3. Частоты встреченных стоп-кодонов

<i>Escherichia coli</i>	<i>Gracilibacteria bacterium</i>	<i>Mycoplasma pneumoniae</i>
TGA 1246 TAA 2761 TAG 306	TGA 1 TAA 1000 TAG 188	TGA 0 TAA 533 TAG 221

В представленных в таблице результатах можно отметить полное отсутствие кодона TGA у организма *Mycoplasma pneumoniae* и аналогичное почти полное отсутствие этого же кодона у *Gracilibacteria bacterium*. Так происходит по причине особенностей генетического кода некоторых прокариот. Так у бактерий составляющих микробиоту человеческого

организма и родственных им таксонов, в том числе у *Gracilibacteria bacteria* кодон TGA кодирует глицин [4]. Аналогично происходит у организмов рода *Mycoplasmatales* [5].

Задание 5: GC - skew

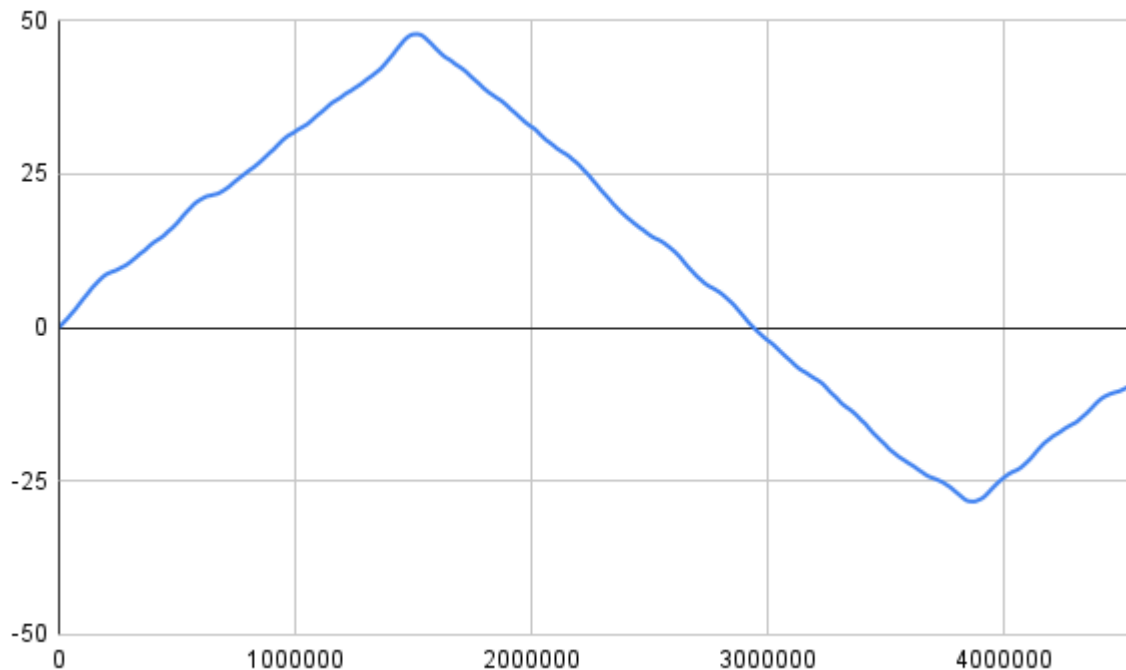


Рисунок 1. График зависимости GC-skew с окном 100000 и шагом 1000 полной последовательности генома *E. Coli*

Данный метод позволяет предсказывать координаты мест начала и окончания транскрипции по локальному изменению нуклеотидного соотношения [6].

По полученным данным максимальное значение cumulative GC-skew (47,733) приходится на участок с координатой 1315000, минимальное (-28,328) соответствует точке 3870000.

Соответственно предположительное расположение ориджина транскрипции в области 1315000 нуклеотида.

Список литературы:

- [1] Xiongwen Cao, Sarah A. Slavoff: Non-AUG start codons: expanding and regulating the small and alternative ORFeome: Exp Cell Res. 2020, June 1
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7256928/>
- [2] Surinder Kumar Indian: *Mycoplasma pneumoniae*: A significant but underrated pathogen in paediatric community-acquired lower respiratory tract infections: Department of Microbiology, Maulana Azad Medical College, J Med Res. Jan
<https://pubmed.ncbi.nlm.nih.gov/29749357/>
- [3] T. C. Stadtman: Selenocysteine: Annu Rev Biochem . 1996
<https://pubmed.ncbi.nlm.nih.gov/8811175/>
- [4] James H Campbell et. al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota: Proc Natl Acad Sci U S A . 2013 Apr
<https://pubmed.ncbi.nlm.nih.gov/23509275/>
- [5] J M Bové 1 Molecular features of mollicutes
Clin Infect Dis 1993 Aug
<https://pubmed.ncbi.nlm.nih.gov/7691196/>
- [6] A Grigoriev, Analyzing genomes with cumulative skew diagrams, Nucleic Acids Res. 1998 May 15
<https://pubmed.ncbi.nlm.nih.gov/9580676/>

Сопроводительные материалы:

1. Тексты скриптов:
https://colab.research.google.com/drive/1J3GuAGrhB9hz9pTNfy2QfQi_rHIRV5Zu?usp=sharing
2. Таблица для построения графика в задании 5:
https://docs.google.com/spreadsheets/d/1xTYI97XC4TwxCxQ3qHJHsyyuwT6_L0757b4VGnmbQ1I/edit?usp=sharing